International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1619-1623 www.ijaem.net ISSN: 2395-5252

Heart Disease Prediction Using Hrflm

Sweatha . M, Ramya.S, Icewariiya.S,
Dr.V.Kalaivani(Professor)
Computer Science and Engineering,National Engineering College, Kovilpatti.

Submitted: 05-06-2022 Revised: 17-06-2022 Accepted: 20-06-2022

ABSTRACT: In recent times, Heart Disease prediction has been one of the most complicated tasks in the medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing vast amounts of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid associated risks and alert the patient well in advance. This article makes use of a heart disease dataset available in the UCI machine learning repository. The proposed work predicts the chances of Heart Disease. It classifies the patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression, and Random Forest. Thus, this project presents a comparative study by analyzing the performance of different machine learning algorithms. The trial results verify that the Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

The heart is one of the most important parts of the body. It helps to purify and circulate blood to all parts of the body. Most deaths in the world are due to Heart Diseases. Some symptoms like chest pain, faster heartbeat, discomfort in breathing are recorded. This data is analysed regularly. In this review, an overview of heart disease and its current procedures is firstly introduced. Furthermore, an in-depth analysis of the most relevant machine learning techniques available in the literature for heart disease prediction is briefly elaborated. The discussed machine learning algorithms are Decision Tree, SVM, ANN, Naive Bayes, Random Forest, KNN. The algorithms are compared based on features. We are working on the algorithm with the best accuracy. This will help the doctors to assist the heart problem easily.

Keywords: Heart Disease, Random Forest , Machine Learning , Linear Regression , Chest pain, Prediction

I. INTRODUCTION

Heart disease is one of the major causes of liver complications and subsequently leads to death. The heart disease diagnosis and treatment are very complex, especially in developing countries, due to the rare availability of efficient diagnostic tools and shortage of medical professionals and other resources which affect proper prediction and treatment of patients. Inadequate preventive measures, lack of experienced or unskilled medical professionals in the field are the leading contributing factors. Although a large proportion of heart diseases are preventable, they continue to rise mainly because preventive measures are inadequate. In today's digital world, several clinical decision support systems on heart disease prediction have been developed by different scholars to simplify and ensure efficient diagnosis. This project investigates the state of the art of various clinical decision support systems for heart disease prediction, proposed by various researchers using data mining and machine learning techniques. Classification algorithms such as the Naive Bayes (NB), Decision Tree (DT), an Artificial Neural Network (ANN) have been widely employed to predict heart diseases, where various accuracies were obtained. Hence, only marginal success is achieved in the creation of such predictive models for heart disease patients. Therefore, there is a need for more complex models that incorporate multiple geographically diverse data sources to increase the accuracy of predicting the early onset of the disease.

Heart disease is the kind of disease that can cause death. Each year too many people are dying due to heart disease. Heart disease can occur due to the weakening of heart muscle. Also, heart failure can be described as the failure of the heart to pump blood. Heart disease is also called coronary artery disease (CAD). CAD can occur due to insufficient blood supply to arteries.

Heart disease can be detected using symptoms like high blood pressure, chest pain, hypertension, cardiac arrest, etc. There are many

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1619-1623 www.ijaem.net ISSN: 2395-5252

types of heart diseases with different types of symptoms. Like: 1) heart disease in blood vessels: chest pain, shortness of breath, pain in neck and throat. 2)heart disease caused by abnormal heartbeats: slow heartbeat, discomfort, chest pain., etc. The most common symptoms are chest pain, shortness of breath, discomfort, chest pain, etc. The most common symptoms are chest pain, shortness of breath, and fainting. Causes of heart disease are defects you're born with, high blood pressure, diabetes, smoking, drugs, alcohol. Sometimes in heart disease, the infection also affects the inner membrane which is identified by symptoms like fever, fatigue, dry cough, skin rashes. Causes of heart infection are bacteria, viruses, parasites. There are several other causes for heart disease that makes the disease worst. It includes consumption of alcohol and drugs and smoking .All this heart disease has to be predicted and treated at early stages .

II. RELATED WORKS

Heart Disease is the most dangerous life-threatening chronic disease globally. The objective of the work is to predict the occurrence of heart disease in a patient using a random forest algorithm. Then it was processed using python open-access software in jupyter notebook. The datasets are classified and processed using the machine learning algorithm Random forest[1] The healthcare sector produces enormous amounts of data every day. This data helps to extract the hidden information, which is useful to predict disease at the earliest. In the medical field, predicting heart disease is treated as one of the intricate tasks. Therefore, there is a necessity to develop a decision support system to forecast the heart disease[2] Heart disease is a leading cause of premature death in the world. Predicting the outcome of the disease is a challenging task. Data mining is involved to automatically infer diagnostic rules and help specialists to make the diagnosis process more reliable. Several data mining techniques are used by researchers to help health care professionals to predict heart disease. Random forest is an ensemble and most accurate learning algorithm, suitable for medical applications.[3] Heart disease is the kind of disease that can cause death. Each year too many people are dying due to heart disease. Heart disease can occur due to the weakening of heart muscle. Also, heart failure can be described as the failure of the heart to pump blood. Heart disease is also called coronary artery disease (CAD). CAD can occur due to insufficient blood supply to arteries.[4] This questionnaire investigated a variety of existing technologies that supported data mining to predict heart disease. In this survey, we learned how to apply data mining techniques to predict heart attacks. Previously, leaving the system was mostly a design using a single algorithm with poor accuracy. In some researchers' studies, we have observed that more accuracy can be achieved through the hybridization of two or more algorithms. [5] Heart disease is one of the major causes of liver complications and subsequently leads to death. The heart disease diagnosis and treatment are very complex, especially in developing countries, due to the rare availability of efficient diagnostic tools and shortage of medical professionals and other resources which affect proper prediction and treatment of patients. Inadequate preventive measures, lack of experienced or unskilled medical professionals in the field are the leading contributing factors.[6] In this paper, we have compared the classification results of two models i.e. Random Forest and the J48 for classifying twenty versatile datasets. We took 20 data sets available from the UCI repository [1] containing instances varying from 148 to 20000. We compared the classification results obtained from methods i.e. Random Forest and Decision Tree . The classification parameters consist of correctly classified instances, incorrectly classified instances, F-Measure, Precision, Accuracy and Recall. We discussed the pros and cons of using these models for large and small data sets.[7] Random forests are a scheme proposed by Leo Breiman in the 2000s for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data. Despite growing interest and practical use, there has been little exploration of the statistical properties of random forests, and little is known about the mathematical forces driving the algorithm. In this paper, we offer an in-depth analysis of a random forests model suggested by Breiman (2004), which is very close to the original algorithm. We show in particular that the procedure is consistent and adapts to sparsity , in the sense that its rate of convergence depends only on the number of strong features and not on how many noise variables are present.[8] Random forests is a statistical- or machine-learning algorithm for prediction. we introduce a corresponding new command, rforest. We overview the random forest algorithm and illustrate its use with two examples: The first example is a classification problem that predicts whether a credit card holder will default on his or her debt. The second example is a regression problem that predicts the log scaled number of shares of online news articles.[9] Ensemble classification is a data mining approach that utilizes

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1619-1623 www.ijaem.net ISSN: 2395-5252

several classifiers that work together to identify the class label for unlabeled instances. Random forest (RF) is an ensemble classification approach that has proved its high accuracy and superiority. With one common goal in mind, RF has recently received considerable attention from the research community to further boost its performance. In this paper, we look at developments of RF from birth to the present. [10]

III. PRE-PROCESSING

Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) is effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 90.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

We use Hybrid Random Forest and Linear Model in the proposed method is to detect heart disease from the raw dataset collected. The voting classifier is used internally for Random Forest and Linear Model. The algorithm which gives better results is considered for prediction. Pre-Processing of data is using 300 patient records. The feature selection is using the attributes of the patient. Classification is done by using the proposed algorithms. The splitting of data is using decision trees. HRFLM algorithm gives the highest accuracy of 90.7%

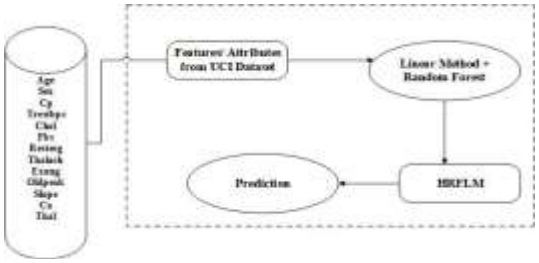

Fig 1 Architecture Diagram

This diagram describes the architecture flow of the process involved in Heart disease prediction. The required features are collected from Dataset and pre-processed. The selected data are now given to a system which is Hybrid Random Forest and Linear Modeling .It is a combination of Random forest and Linear Regression algorithms to predict the Heart disease with high accuracy. This system can predict the heart disease with more than 90% accuracy.

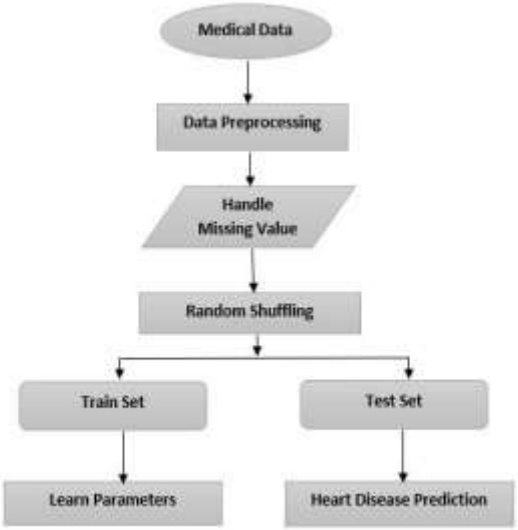

Fig 2 Data Flow Diagram

The process of heart disease prediction can be done as per the Fig 2.The medical data are collected from the Dataset. The data is now preprocessed. Preprocessing of data is nothing but the data set is checked for null values. If any null values are present in the data set then it is dropped. The dataset is now divided into training and test data set. From the training data set the attributes that are highly responsible for predicting Heart disease are learnt. From the Testing data set Heart disease is now predicted.

IV. METHODOLOGY

4.1. DATA PRE-PROCESSING

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing. The multiclass variable and binary classification are introduced for the attributes of the given dataset. The multi-class variable is used to check the presence or absence of heart disease. In the instance of the patient having heart disease, the value is set

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1619-1623 www.ijaem.net ISSN: 2395-5252

to 1, else the value is set to 0 indicating the absence of heart disease in the patient. The pre-processing of data is carried out by converting medical records into diagnosis values. The results of data pre-processing for 297 patient records indicate that 137 records show the value of 1 establishing the presence of heart disease while the remaining 160 reflected the value of 0 indicating the absence of heart disease.

4.2. FEATURE SELECTION AND REDUCTION

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease. The experiment was done with the ML technique namely using all 13 attributes.

4.3.CLASSIFICATION MODELLING

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error. The performance is further optimized by choosing the DT cluster with a high rate of error and extraction of its corresponding classifier features. The performance of the classifier is evaluated for error optimization on this data set

4.4. DISEASE PREDICTION USING HRFLM

Once the classification is done now heart disease has to be predicted. For the prediction of heart disease the patient data is given into the windows application. The data is given as a file that contains important attributes needed for prediction.

V. RESULT

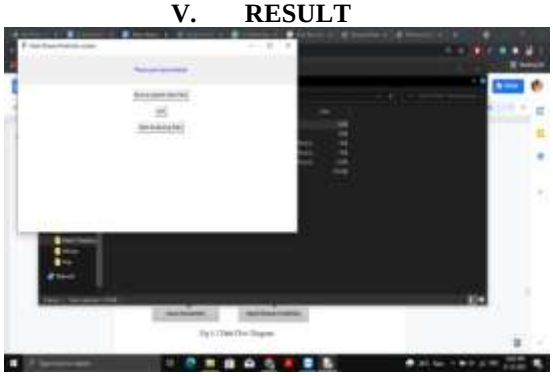

Fig 3 Windows Application for Heart Disease Prediction System using HRFLM

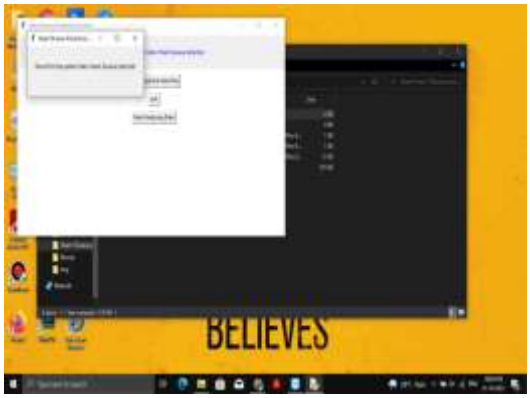

Fig 4 Prediction of Heart Disease – Detected

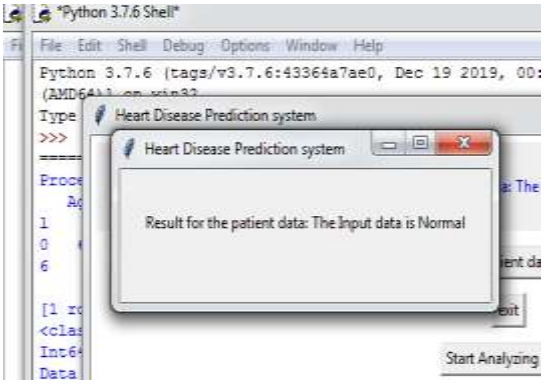

Fig 5 Prediction of Heart Disease – Not Detected

VI. CONCLUSION

Out of the 13 features we examined, the top 4 significant features that helped us classify between a positive & negative Diagnosis were chest pain type (cp), maximum heart rate achieved (thalach), number of major vessels (ca), and ST depression induced by exercise relative to rest (oldpeak).Our machine learning algorithm can now classify patients with Heart Disease. Now we can properly diagnose patients, & get them the help they need to recover. By detecting these features early, we may prevent worse symptoms from arising later. Our Hybrid Random Forest and Linear model algorithm yields the highest accuracy, 80%. Any accuracy above 70% is considered good.

VII. FUTURE SCOPE

This project can be further improved in future by adding constraints for predicting other diseases like Brain tumour and other life risking diseases. This project can be further improved by applying other combinations of Machine Learning Algorithms that can have high accuracy rate.

REFERENCES

[1]. Madhumita Pal, Smita Parija,Prediction of Heart Diseases using Random Forest, J. Phys.: Conf. Ser. 1817 012009
[2]. Malavika G, Rajathi N, Vanitha V and Parameswari P,Heart Disease Prediction Using Machine Learning Algorithms,Vol. 4 No. 2(112) (2021): Information technology. Industry control systems
[3]. M.A.Jabbar1, B.L.Deekshatulu2 and Priti Chandra, Intelligent heart disease prediction system using random forest and evolutionary approach,Journal of Network and Innovative Computing ISSN 2160-2174 Volume 4 (2016)
[4]. Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, Pushkar Patil,A Survey on Prediction Techniques of Heart Disease using Machine Learning,(IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS060298 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 9 Issue 06, June-2020
[5]. Apruv Patel, kunjan D. Khatri, Smit Kiri, Kathan Patel,A Literature Review on Heart Disease Prediction Based on Data Mining Algorithm, 2018 IJRTI | Volume 3, Issue 6 |
[6]. Lamido Yahaya,Etemi Garba, A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques,American Journal of Artificial Intelligence 2020; 4(1): 20-29
[7]. Jehad Ali,Rehanullah Khan,Nasir Ahmad, Random Forests and Decision Trees,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
[8]. Gerard Biau, Analysis of a Random Forests Model,Journal of Machine Learning Research 13 (2012) 1063-1095
[9]. Matthias Schonlau, The random forest algorithm for statistical learning,The Stata Journal (2020) 20, Number 1, pp. 3–29
[10]. Khaled Fawaregh, Mohamed Medhat Gaber,Random forests: from early developments to recent advancements,Systems Science & Control Engineering: An Open Access Journal, 2014